VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

WINTER SEMESTER 2022-2023

HUMAN COMPUTER INTERACTION

COURSE CODE- CSE4015

Project Report

ASL Translation

Submitted to : Prof. Dr. S M Satapathy

Team Members :

Harshil Kothari 20BCE0152
Rishabh Agrawal 20BCE0798
Zaman Saleel 20BCE2025
Karthik S Menon 20BCE2555

Table of Contents

Project Overview (Abstract)

The communication gap between hearing and non-hearing individuals has been a significant challenge in society, and American Sign Language (ASL) is a convenient form of communication primarily used by the hard-of-hearing or deaf. However, the lack of knowledge and experience with ASL in the global population has been a major issue.
This is where real-time sign language detection aid can be a game-changer, helping to bridge the communication gap between hearing and non-hearing individuals. The technology of text and speech to ASL conversion, and vice versa, has the potential to provide an accessible and convenient way for deaf and hard-of-hearing individuals to communicate with hearing individuals and access spoken information.
The objective of text-to-ASL conversion is to enable deaf and hard-of-hearing individuals to receive information that is accessible and understandable to them, while speech-to-ASL conversion aims to provide a way for them to access spoken information and communicate with hearing individuals.
ASL to text and speech conversion technologies also aim to provide mute individuals with an accessible way to communicate with others, whether it is through written text or spoken language. Moreover, combining most of the common ways of communication in society and allowing deaf and mute individuals to overcome a variety of their difficulties of day-to-day life is a significant benefit of ASL identification and translation. The use of computer vision algorithms, machine learning models, automatic speech recognition (ASR), and text-to-speech (TTS) systems enable the development of such technologies.
In this context, our project focuses on the use of OpenCV to capture ASL gestures, which are then passed through a CNN model for prediction, followed by NLP processing of the outputted English sentences. The project uses CNN models that contain four convolution layers and two fully connected Dense layers, which are trained on the ASL dataset. The dropout of 0.25 in each of the layers of the CNN ensures that there is no overfitting of the model on the training data.
The development of text and speech to ASL conversion technologies can make a significant contribution to bridging the communication gap between hearing and non-hearing individuals. The real-time sign language detection aid can be

particularly useful while video-conferencing and during social interactions, enabling access to spoken information and communication with hearing individuals. The ASL identification and translation technology can have a significant impact on the lives of deaf and mute individuals by enabling them to use HCI to communicate with the rest of society.

Introduction

Text and speech to American Sign Language (ASL) conversion, and vice versa, refers to the process of converting written text or spoken language into ASL gestures and signs, and converting ASL gestures and signs into written text or spoken language. These technologies are typically achieved through the use of computer vision algorithms, machine learning models, automatic speech recognition (ASR), and text-to-speech (TTS) systems. The goal of these technologies is to bridge the communication gap between hearing and non-hearing individuals and to provide deaf and hard-of-hearing individuals with an accessible and convenient way to communicate with hearing individuals and to access spoken information. Additionally, these technologies can be useful for individuals who are learning ASL, as they provide an easy way to practice and review the signs and gestures.

Background / Related Work

No.	Title	Findings	Authors
1	A machine learning based approach for the detection and recognition of Bangla sign language	It demonstrates Hand Gesture recognition which is performed using HOG (Histogram of Oriented Gradients) for extraction of features from the gesture image and SVM (Support Vector Machine) as classifier. Finally, predict the gesture image with output text. This output text is converted into audible sound using TTS (Text to Speech) converter.	Muttaki Hasan Tanvir Hossain Sajib Mrinmoy Dey
2	Speech Recognition Automation by ASR	Author has presented multiple experiments to design a statistical model for deaf people for the conversion to sign language from the speech set. They have further made the system that automates the speech recognition by ASR by the help of animated demonstration and translation statistical module for multiple sets of signs. This paper demonstrates the process that translates the speech by automation recognizer having all three mentioned configurations. The paper came up with the result with finite type state transducer having the word error rate among the range of 28.21% and 29.27% for the output of ASR.	Matthew Zajechowski

3	Vision-based sign language translation device	The proposed system which is an interactive application program developed using LABVIEW software and incorporated into a mobile phone. The sign language gesture images are acquired using the inbuilt camera of the mobile phone; vision analysis functions are performed in the operating system and provide speech output through the inbuilt audio device thereby minimizing hardware requirements and expense. The experienced lag time between the sign language and the translation is little because of parallel processing. This allows for almost instantaneous recognition from finger and hand movements to translation. This is able to recognize one handed sign representations of alphabets (A-Z) and numbers (0-9). The results are found to be highly consistent, reproducible, with fairly high precision and accuracy	Yellapu Madhuri, G Anitha, M Anburajan
4	Indian Sign Language (ISL) Translation System For Sign Language Learning	It functions continuously by offering a sequence of sign language gestures to create an automated training set and by offering the spots signs from the set of training. They have put forth a system that supervises the sentence and determines the associated compound sign gesture using a supervision of noisy texts, using instance learning as a density matrix technique. The group that was first intended to demonstrate the continuous data stream of words is now used as a training group for identifying gesture	M.Jerin Jose, V. Priyadarshini, M.Suresh Anand, A.Kumaresan, Dr.N. Mohan Kumar

		posture. They have experimented with this small sample of automated data that is used for their training, identification, and storage of subtle sign data.	
5	ML Based Sign Language Recognition System	It reviews different steps in an automated sign language recognition (SLR) system. The model is based on vision-based isolated hand gesture detection and recognition. The model made use of a convex hull for feature extraction and KNN for classification.	A.Adeyanju, O.O.Bello, M.A.Adegboye
6	A real-time portable sign language translation system	It uses the wireless system to process the data. To differentiate hand motion, they have inner sensors put into gloves to show the parameters as given by, posture, orientation, motion, defined of the hand in Taiwanese Sign Language could be recognize in no error. The hand gesture is considered by flex inner sensor and the palm size considered using the g sensor and the movement is considered using the gyroscope. Input signals would have to be consider for testing for the sign to be legal or not periodically. As the signal which was sampled can stay longer than the pre-set time, the legal gesture sent using phone via connectivity like Bluetooth for differentiating gestures and translates it. With the proposed architecture and algorithm, the accuracy for gesture recognition is quite satisfactory. As demonstrated the result get the accuracy of 94% with the concurrent architecture.	Lih-Jen Kau, Wan-Lin Su, Pei-Ju Yu, Sin-Jhan Wei

7	A survey on 3D hand pose estimation: Cameras, methods, and datasets	This paper contains a comprehensive survey, including depth cameras, hand pose estimation methods, and public benchmark datasets. First, a markerless approach is proposed to evaluate the tracking accuracy of depth cameras with the aid of a numerical control linear motion guide. Second, the methods and lines of research are summarized. Third, existing benchmark datasets and evaluation criteria are identified to provide further insight into the field of hand pose estimation.	Li Rui, Liu Zhenyu, Tan Jianrong
8	Vision-based hand pose estimation: A review	A review on various CV based hand pose estimations were done. CV has a distinctive role in the development of direct sensing-based HCI. However, various challenges must be addressed in order to satisfy the demands of potential interaction methods. Currently, CV-based pose estimation has some limitations in processing arbitrary hand actions. Incorporating the full functionality of the hand in HCI requires capturing the whole hand motion. However, CV can only provide support for only a small range of hand actions under restrictive conditions.	Ali Erol. George Bebis, Mircea Nicolescu, Richard D. Boyle, Xander Twombly
9	Facial expressions recognition for arabic sign language translation	Arabic Sign Language (ArSL) tends to be a descriptive gesture language, facial expressions are involved in 70% of total signs.The system employed already existing technical methods such as: Recursive Principle Components (RPCA) for feature	A.S. Elons; Menna Ahmed; Hwaidaa Shedid

		extraction and Multi-layer Perceptron (MLP) for classification.	
10	Automated Sign Language Interpreter	It demonstrates Instrumented gloves with audio out are the solution here. The gloves attached with various sensors are worn for sign interpretation. Hence, the proposed system solves the problem and helps the dumb people in communication with the rest of the world at low cost.	Hardik Rewari, Vishal Dixit, Dhroov Batra, Hema N
11	Sign language interpreter using a smart glove	It demonstrates a novel approach of interpreting the sign language using the portable smart glove. LED-LDR pair on each finger senses the signing gesture and couples the analog voltage to the microcontroller	Nikhita Praveen Naveen Karanth M S Megha
12	Automatic detection of learners effect from gross body language.	We explored the reliability of detecting learners' affect by monitoring their gross body language body position and arousal	Sidney D’Mello, and art Graesser.
13	A Survey on Hand Pose Estimation with Wearable Sensors and Computer-Vision-Based Methods	The purpose of this survey is to conduct a comprehensive and timely review of recent research advances in sensor-based hand pose estimation, including wearable and vision-based solutions. Hand kinematic models are firstly discussed. An in-depth review is conducted on data gloves and vision-based sensor systems with corresponding modeling methods. Particularly, this review also discusses deep-learning-based methods,	Weiya Chen ,Chenchen Yu, Chenyu Tu, Zehua Lyu, Jing Tang, Shiqi Ou, Yan Fu and Zhidong Xue

		which are very promising in hand pose estimation. Moreover, the advantages and drawbacks of the current hand gesture estimation methods, the applicative scope, and related challenges are also discussed.	
14	Hand pose estimation and tracking in real and virtual interaction:A review	The goal of this survey is to develop an up-to-date taxonomy of the state-of-the-art vision-based hand pose estimation and tracking methods with a new classification scheme: hand-object interaction constraints. This taxonomy allows us to examine the strengths and weaknesses of the current state of the art and to highlight future trends in the domain.	Ammar Ahmad, Cyrille Migniot, Albert Dipanda
15	Trajectory-based recognition of dynamic Persian sign language using hidden Markov model	Existing Persian sign language recognition systems are mainly restricted to static signs which are not very useful in everyday communications.These time-varying trajectories were then modeled using Hidden Markov Model (HMM) with Gaussian probability density functions as observations.	Saeideh GhanbariAzar, Hadi Seyedarabi

Real-life Applicability

This project of converting text, speech and video to American Sign Language (ASL) and vice versa has significant real-life applicability, particularly for individuals who are deaf, hard-of-hearing, or mute.

One potential application of this project is in education. Many schools for the deaf or hard-of-hearing use ASL as the primary mode of communication. With the help of this project, teachers and students can easily convert text and speech into ASL, making it easier for students to understand and follow the lessons. Additionally, this technology can also be useful for individuals who are learning ASL, as it provides an easy way to practice and review the signs and gestures they have learned.

Another potential application of this project is in the workplace. With the increasing use of video conferencing, individuals who are deaf, hard-of-hearing, or mute may have difficulty communicating with their colleagues. The real-time sign language detection aid provided by this project can solve this issue by translating the sign language into spoken or written language in real-time, enabling everyone to understand and participate in the conversation.

This technology can also be useful for medical professionals, police officers, and emergency responders who may encounter individuals who are deaf or hard-of-hearing during their work. By providing a way to communicate with these individuals using ASL, these professionals can ensure that they receive the same level of care and attention as others.

Overall, this project has the potential to make a significant impact on the lives of individuals who are deaf, hard-of-hearing, or mute, by providing them with a more accessible and convenient way to communicate with others and access information.

Individual Contributions

Harshil Kothari	-	New Frontend
Rishabh Agrawal	-	Asl to Speech
		Live Video Capture and Frame Extraction
		Asl to Text
Zaman Saleel	-	Text to Asl Backend
		Old Frontend
Karthik S Menon	-	Text to Asl Frontend
		Old Frontend

Tools and Technologies used / Hardware – Software Requirements

Libraries: Numpy, Pandas, TensorFlow, Keras, matplotlib

APKs: Google Text to Speech

Dataset: Kaggle - https://www.kaggle.com/datasets/grassknoted/asl-alphabet

Proposed System Process Flow (Pictorial Representation)

Working Methodology

We have used OpenCV to capture images of the gestures to be translated from the system's video/camera input. These images are adjusted to the appropriate size (128x128 RGB) and passed to our CNN model for prediction.

The CNN contains 4 convolution layers using ReLU as the activation function and Max Pooling. These layers are followed by a flattening layer and two fully connected Dense layers of neurons with ReLU activation. Finally, the output layer is a dense layer containing 29 neurons to represent all the possible classes of our dataset. The output layer uses a Softmax activation function. We have also applied a dropout of 0.25 in each of the layers of our CNN to ensure there is no overfitting of our model on the training data. The resulting CNN has a validation accuracy of 99%

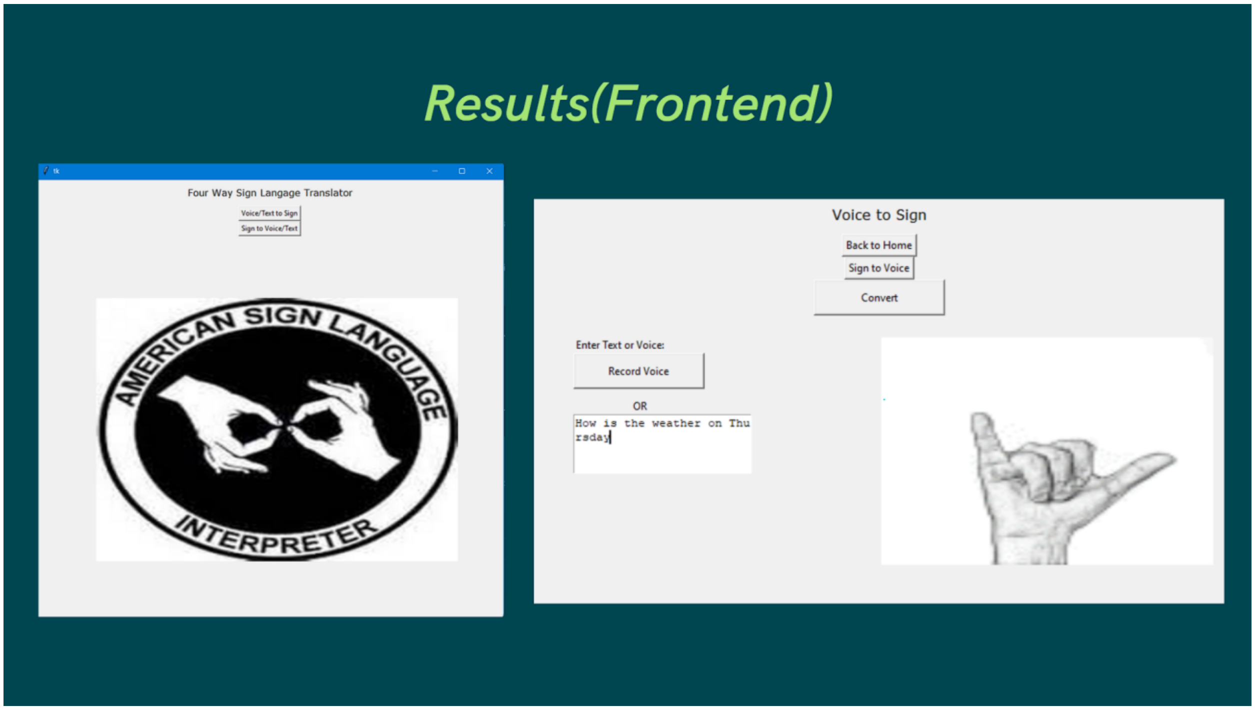
```
Epoch 20: accuracy improved from 0.94228 to 0.94576, saving model to model_weights1.h5
2175/2175 [==============================] - 897s 412ms/step - loss: 0.1663 - accuracy: 0.9458 - val_loss: 0.0236 - val_accuracy: 0.9924

Done.
544/544 [==============================] - 24s 42ms/step - loss: 0.0236 - accuracy: 0.9924
[0.023579731583595276, 0.9924138188362122]
```

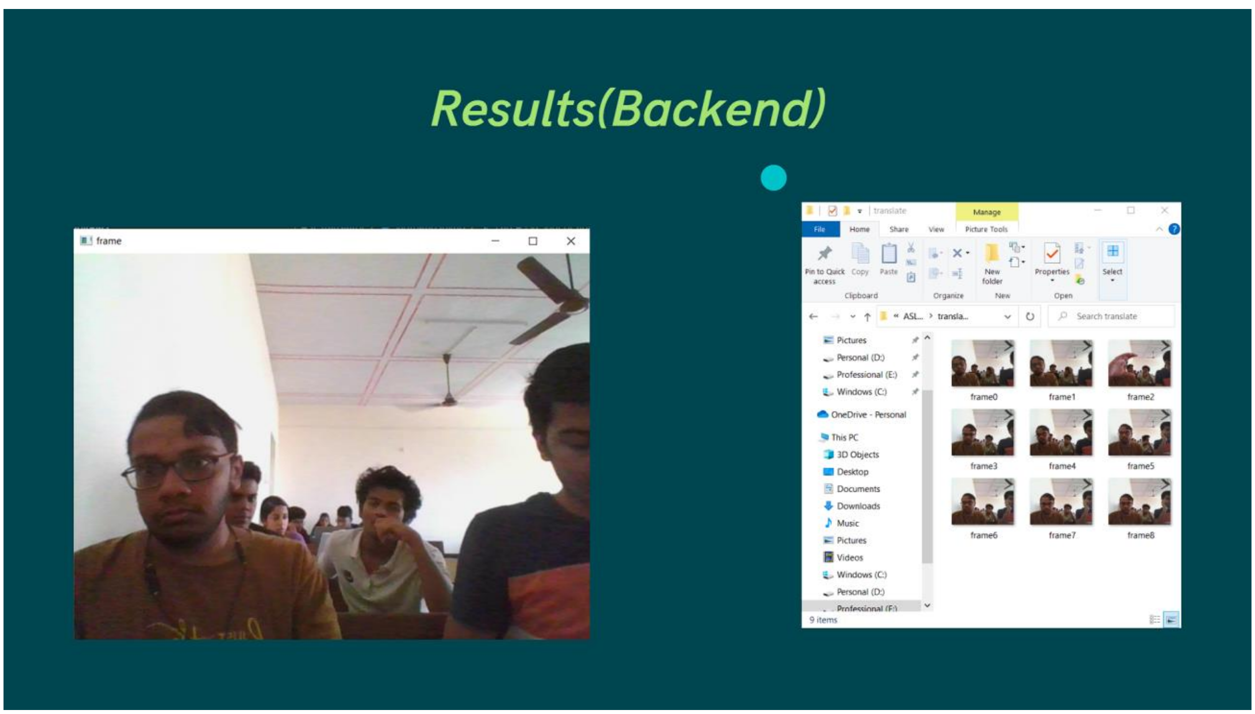
The images after passing through the CNN give us an output with the predicted class (Translated ASL). This character is added onto our translated string. The translated string, at the end of a sentence, is passed through an NLP model which identifies the words within and inserts the spaces required to turn it into a proper sentence.

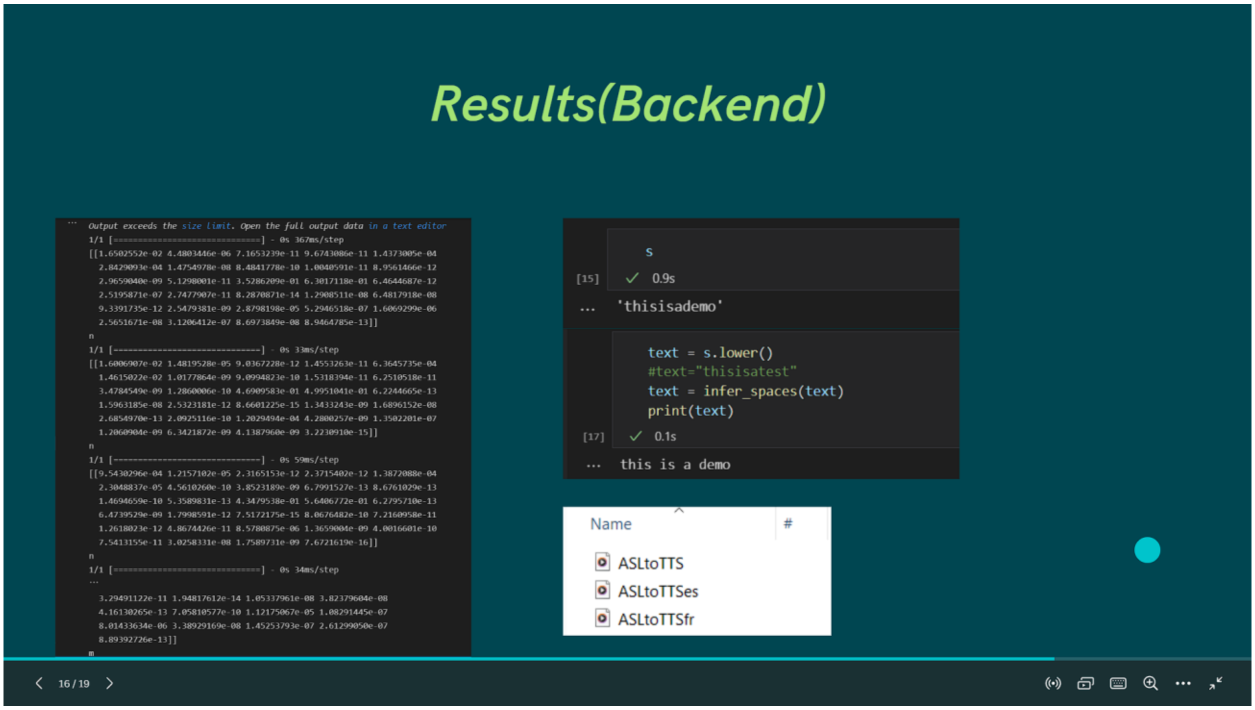
Once the ASL gestures have been translated into a proper English sentence, they can be outputted either as plain text or as speech with the use of Google's speech to text API.

For translating normal text into ASL, we just have to query each character of the text in our database and display the associated ASL gesture image.

Implementation Results and User Interfaces






Results(Backend)
Output exceeds the size limit. Open the full output data in a text editor
1/1 [==============================] - 0s 367ms/step
[[1.6502552e-02 4.4803446e-06 7.1653239e-11 9.6743086e-11 1.4373005e-04
2.8429093e-04 1.4754978e-08 8.4841778e-10 1.0040591e-11 8.9561466e-12
2.9659040e-09 5.1298001e-11 3.5286209e-01 6.3017118e-01 6.4644687e-12
2.5195871e-07 2.7477907e-11 8.2870871e-14 1.2908511e-08 6.4817918e-08
9.3391735e-12 2.5479381e-09 2.8798198e-05 5.2946518e-07 1.6069299e-06
2.5651671e-08 3.1206412e-07 8.6973849e-08 8.9464785e-13]]
n
1/1 [==============================] - 0s 33ms/step
[[1.6006907e-02 1.4819528e-05 9.0367228e-12 1.4553263e-11 6.3645735e-04
1.4615022e-02 1.0177864e-09 9.0994823e-10 1.5318394e-11 6.2510518e-11
3.4784549e-09 1.2860006e-10 4.6909583e-01 4.9951041e-01 6.2244665e-13
1.5963185e-08 2.5323181e-12 8.6601225e-15 1.3433243e-09 1.6896152e-08
2.6854970e-13 2.0925116e-10 1.2029494e-04 4.2800257e-09 1.3502201e-07
1.2060904e-09 6.3421872e-09 4.1387960e-09 3.2230910e-15]]
n
1/1 [==============================] - 0s 59ms/step
[[9.5430296e-04 1.2157102e-05 2.3165153e-12 2.3715402e-12 1.3872088e-04
2.3048837e-05 4.5610260e-10 3.8523189e-09 6.7991527e-13 8.6761029e-13
1.4694659e-10 5.3589831e-13 4.3479538e-01 5.6406772e-01 6.2795710e-13
6.4739529e-09 1.7998591e-12 7.5172175e-15 8.0676482e-10 7.2160958e-11
1.2618023e-12 4.8674426e-11 8.5780875e-06 1.3659004e-09 4.0016601e-10
7.5413155e-11 3.0258331e-08 1.7589731e-09 7.6721619e-16]]
n
1/1 [==============================] - 0s 34ms/step
...
3.29491122e-11 1.94817612e-14 1.05337961e-08 3.82379604e-08
4.16130265e-13 7.05810577e-10 1.12175067e-05 1.08291445e-07
8.01433634e-06 3.38929169e-08 1.45253793e-07 2.61299050e-07
8.89392726e-13]]
m
s
[15] 0.9s
... 'thisisademo'
text = s.lower()
#text="thisisatest"
text = infer_spaces(text)
print(text)
[17] 0.1s
... this is a demo
Name
#
ASLtoTTS
ASLtoTTSes
ASLtoTTSfr
16 / 19

Interfaces Validation with Nielsen’s 10 point heuristics

i. **Visibility of system status:** Our interface satisfies this heuristic as it provides feedback to the user throughout the various stages of each functionality, such as displaying the progress of video processing.

ii. **Match between system and the real world:** Our interface satisfies this heuristic as it uses simple and easy-to-understand language and concepts that are familiar to the user.

iii. **User control and freedom:** Our interface satisfies this heuristic by providing an "emergency exit" to cancel or undo actions and allowing users to easily navigate between functionalities.

iv. **Consistency and standards:** Our interface satisfies this heuristic by maintaining consistency in the layout and design across all functionalities.

v. **Error prevention:** Our interface satisfies this heuristic by providing clear instructions and implementing validation checks to prevent user errors.

vi. **Recognition rather than recall:** Our interface satisfies this heuristic by providing visual cues for the different functionalities and using easily recognizable icons and symbols.

vii. **Flexibility and efficiency of use:** Our interface satisfies this heuristic by providing keyboard shortcuts and allowing users to adjust the settings for the video-to-image processing.

viii. **Aesthetic and minimalist design:** Our interface satisfies this heuristic by keeping the interface visually appealing and only showing the necessary information for each functionality.

ix. **Help users recognize, diagnose, and recover from errors:** Our interface satisfies this heuristic by providing clear and specific error messages that guide users towards recovery options.

x. **Help and documentation:** Our interface satisfies this heuristic by providing access to documentation and help resources to explain how to use the different functionalities.

Comparative Analysis (with other existing technologies discussed in Related Works)

Communication Accessibility: Traditional ASL relies on face-to-face interaction between sign language users, which can be challenging when communicating with non-sign language users who do not understand ASL. In contrast, a two-way sign language translator can bridge this communication gap by translating sign language into spoken or written language, making communication more accessible to a wider audience.

Real-time Translation: One of the key advantages of a two-way sign language translator is its ability to provide real-time translation. This allows for faster and more efficient communication compared to traditional ASL, which may require additional time for interpreting and translation between sign language and spoken or written language.

Accuracy and Consistency: While traditional ASL interpretation can vary depending on the skills and fluency of the interpreter, a two-way sign language translator can provide consistent and accurate translations, eliminating potential inconsistencies and misunderstandings that may arise in traditional ASL communication.

Portability and Convenience: A two-way sign language translator can be a portable device or software installed on a mobile device, providing convenience and flexibility for sign language users to communicate in various settings and situations. Traditional ASL interpretation, on the other hand, may require the presence of an interpreter, which may not always be feasible or convenient.

User Experience: Traditional ASL relies on physical gestures, facial expressions, and body language for communication, which can convey rich nuances of meaning. While a two-way sign language translator can provide basic translation, it may not fully capture the nuances of ASL and may lack the expressiveness of traditional ASL communication.

Cost: Traditional ASL interpretation typically involves hiring a qualified interpreter, which can be costly depending on the availability and demand for interpreters. In contrast, a two-way sign language translator can be a one-time investment in a device or software, potentially providing cost-effective communication options in the long run.

Technical Limitations: Two-way sign language translators may have limitations in accurately translating regional variations of sign languages or complex ASL expressions that rely on spatial grammar, classifiers, and non-manual markers. Traditional ASL interpretation, on the other hand, may be more flexible and adaptable to different signing styles and variations.

In summary, while traditional ASL is a natural and expressive mode of communication for sign language users, a two-way sign language translator can provide additional accessibility, convenience, and efficiency in communicating with non-sign language users.

Conclusion & Future Scope

The project is a simple demonstration of how CNN can be used to solve computer vision problems with an extremely high degree of accuracy. The project can be extended to other sign languages by building the corresponding dataset and training the CNN. While Sign languages rely more on context rather than spelling out the words, the paper is able to solve a subset of the Sign Language translation problem. The main objective has been achieved, that is, the need for an interpreter has been eliminated.

The Future Scope of the project includes increasing the accuracy, adding more signs, bug fixes if any and incorporating more sign languages.

References

i. https://ieeexplore.ieee.org/document/7282137
ii. https://www.academia.edu/67849732/Indian_Sign_Language_ISL_Translation_System_For_Sign_Language_Learning
iii. https://www.researchgate.net/publication/328912139_Automated_Sign_Language_Interpreter
iv. https://ieeexplore.ieee.org/abstract/document/7835387
v. https://ieeexplore.ieee.org/iel7/6991454/7002373/07002401.pdf
vi. https://usabilitygeek.com/automatic-speech-recognition-asr-software-an-introduction/
vii. https://ieeexplore.ieee.org/abstract/document/6508395
viii. https://www.sciencedirect.com/science/article/abs/pii/S0031320319301724
ix. https://www.mdpi.com/1424-8220/20/4/1074
x. https://www.sciencedirect.com/science/article/abs/pii/S1077314206002281
xi. https://www.sciencedirect.com/science/article/abs/pii/S0262885619300861
xii. https://ieeexplore.ieee.org/document/7030980
xiii. https://ieeexplore.ieee.org/document/9399594
xiv. https://www.researchgate.net/publication/337756755_Trajectory-Based_Recognition_of_Dynamic_Persian_Sign_Language_Using_Hidden_Markov_Model
xv. https://www.researchgate.net/publication/220356141_Automatic_Detection_of_Learner's_Affect_from_Gross_Body_Language

Appendix:

Link to PPT
https://www.canva.com/design/DAFXPctHEuE/lNujQre1ulQ0Y3z5YBCtWg/edit?utm_content=DAFXPctHEuE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Link to Pre-recorded Demonstration video
https://drive.google.com/drive/folders/1eKjTRlBsW2rev476rB6d32fEBWD6PPRA?usp=sharing

Link to access source files and file containing steps to execute the project
https://github.com/EphemeralAnarchist/ASL_Translator/tree/main/V2

Link to Drive Folder containing the trained model
https://drive.google.com/file/d/1Dx8qA3jhPp7yc7BQ-NJAoKACd4HH2zsH/view?usp=sharing